THE OCCUPATIONAL FEMINIZATION OF WAGES

John T. Addison
Durham University Business School, University of South Carolina, and IZA Bonn

Orgul D. Ozturk
University of South Carolina

Si Wang
Hunan University

Abstract

This paper updates the study by Macpherson and Hirsch (1995) of the effects of occupational gender composition on earnings using monthly CPS data for 1973-1993. In the updating process, we correct for biases in this dataset due to the inclusion of imputed earners and the misreporting of occupation. CPS data for 1996-2010 are used to provide cross-sectional estimates of the impact of feminization on wages as well as its contribution to the gender wage gap. Longitudinal CPS data indicates that the negative effects of gender composition on earnings observed in cross section are lessened (much reduced) when we control for observed (unobserved) heterogeneity. These findings are confirmed using much longer panels from the NLSY. Finally, constructing synthetic panels of aging cohorts suggests that wage penalties are largest for younger cohorts in *female* occupations. (134 words)

JEL Classification: J31, J71

Keywords: wages, gender composition, gender wage gap

Corresponding author:
John T. Addison
Department of Economics
Darla Moore School of Business
University of South Carolina
1014 Greene Street
Columbia, SC 29208
USA
E-mail: ecceaddi@moore.sc.edu

At first blush, the negative effect of the gender composition of an occupation – that is, its proportion female – on the earnings of women (and men) in that occupation might seem to offer a blueprint for policy initiatives. Such measures might include quotas to increase the female component in male jobs and comparable worth, together with more conventional instruments seeking to strengthen human capital endowments. Unfortunately, the empirical consensus does not extend much beyond agreement on the stylized facts of earnings disparities that are increasing in feminization. Acting to blunt policy activism, therefore, is disputation as regards the size and persistence of the negative correlation between gender composition and wages as well as disagreement as to underlying causation.

At root, the controversy has a basis in a literature often containing scant controls for observables such as occupational characteristics that may influence earnings and earning development. Further, the number of studies using longitudinal analysis is still somewhat meager. In the latter context the gender composition variable may in practice be correlated with unmeasured skill and taste differences among workers, and in the former with controls for occupational attributes that might reasonably be expected to influence earnings.

The present paper is motivated by an important study of occupational sex segregation by Macpherson and Hirsch (1995) that is notable in three principal respects: first, in its use of several large datasets; second, in its deployment of arguments not typically found in the literature; and, third, in offering a formal longitudinal analysis of wage change. In short, Macpherson and Hirsch investigate whether the material gender composition effects often reported in the literature are real or instead a chimera reflecting occupational characteristics and differences in labor quality and tastes (and other unobserved characteristics) that are correlated with the proportion female in an occupation. If the latter association dominates, the issue then turns on why predominantly female occupations have become associated with job characteristics and worker endowments leading to lower pay.

But Macpherson and Hirsch's analysis ends in 1993, leaving open the question of whether the trends they identify still characterize the data. Furthermore, there is the important question of whether the biases inherent in the main data set largely resulting from the inclusion of imputed earners – which were less of a problem prior to 1993 and indeed only recognized as such in 2004 – are no longer ignorable. Finally, since their findings after controlling for unobserved heterogeneity might be a function of the short panel employed, we also investigate another dataset offering much longer panels to validate or otherwise the results on the importance of unmeasured worker skill or taste differences to reductions in gender composition effects.

At all times the innovations introduced here are key to precision in answering the research questions: does feminization still hurt workers and do women's jobs pay less? The importance of the issue at a time when women constitute around one-half the workforce is the magnitude of the implied waste in resources and of course the design of policies to ameliorate the situation. And even if the direct effect of feminization emerges as modest, calling into question the scope for certain types of policy (e.g. comparable worth), the indirect effect of feminization as a proxy for unmeasured skills, preferences, and job attributes does nothing to discourage research into gender differences. Indeed, it should redirect research resources toward establishing the manner in which the labor market sorts the genders into jobs with different characteristics and productivities and/or the sources of the gender wage gap other than feminization.

The plan of the paper is as follows. We first present cross-sectional estimates of the relation between proportion female in an occupation and wages, paying close attention to the role of occupational skills and job characteristics. Results are provided by year and also for the pooled data set to examine specification differences in the effects of feminization across alternative groups of workers, inter al. A decomposition of the gender wage gap by broad specification and year then assesses the contribution of feminization to the explained and unexplained gaps. The final stage of the analysis controls for unobserved fixed effects in measuring the relation between gender composition and wages. While the longitudinal capacity of the CPS is again used for this purpose, given that matched worker pairs are potentially available only for adjacent years, the CPS panel analysis is supplemented using information from a longer panel, namely the National Longitudinal Survey of Youth 1979. A summary addresses the main findings and their policy implications.

Theoretical Considerations

There are two main explanations for the covariation of wages and the gender composition of occupations. One is human capital theory and the other is discrimination resulting in crowding and possibly to the undervaluation of women's work. Human capital theory is based on choice (Becker, 1985). From a human capital perspective, predominantly male occupations pay more than predominantly female occupations because individuals in the former have chosen to invest more in human capital. Similarly, by reason of their (historically) weaker labor force attachment, women are viewed as choosing occupations in which their skills will depreciate less rapidly during spells of absence from the labor market (Polachek, 1981, 1985).[1] According to the theory of occupational crowding, however, male jobs pay more because women excluded from them by discrimination are shunted into other occupations with no or lesser discrimination and the resulting increased supply of labor (or crowding) lowers their wages (Bergmann, 1974). The caveat is of course that where women are crowded into particular occupations by reason of their preferences, the negative effect of greater feminization may be a costly compensating differential. It may also be the case that persons employed in female-dominated occupations receive lower returns to occupational characteristics (e.g. specific vocational preparation) because their work – so-called "women's work" – is undervalued (Gerhart and El Cheikh, 1991) even though in principle their incumbents are equally well qualified. There is an extensive literature suggesting that wage inequality is socially constructed and that work in women's occupations is undervalued by reason of institutionalized bias against women (see, for example, Treiman and Hartmann, 1981; Kilbourne et al., 1994; Magnusson, 2009) even if the skills required for lower-paid female dominated jobs are comparable to those in better-paid male-dominated jobs. One of the more transparent aspects of bias is the devaluation of caring and nurturing skills associated with women (Hirsch and Manzella, 2015).[2]

Not surprisingly perhaps the standard models are thin on the details of allocation – in short, how individuals progress through a jobs hierarchy. By analogy with the above narrative, this would on the one hand involve consideration of how individuals control those prospects

---

[1]This approach includes notions of firm-specific human capital formation that have been introduced into the occupational sex-segregation literature by Tam (1997) as part of a differential levels-of-specialization approach to pay differences as opposed to occupational sex segregation. On the operationalization and further evaluation of this specialized human capital argument, see Perales (2013).

[2]A separate although related theme is provided by socialization and domestic labor supply theories; see, for example, Clausen (1968), Marini and Brinton (1984), Hakim (2000). And on the role of a gender gap in workplace authority as a potential explanation for occupational feminization and wages, see England et al. (1994).

through the acquisition of knowledge and skills. On the other, it would also encompass the institutionalist challenge based on notions of *social technology* (Osterman, 1987), having to do with the manner in which jobs are structured, the selection of individuals into those jobs, and the valuation of jobs. The present treatment will eschew consideration of the promotion process, despite its potential importance in producing female-dominated and male-dominated jobs and the application of bureaucratic processes, customs, and notions of fairness that may lead to the systematic undervaluation of women's work (e.g. McArthur, 1985).

Nevertheless, the emphasis of the present treatment on measurement – in particular its use of large data sets, detailed occupational controls, and accounting for unobserved productivity differences – should effectively determine the extent of the direct effect of gender composition upon wages and assist in policy design as well as addressing the reach of broad-based theoretical explanations. Thence, more detailed investigation of, say, the manner in which the labor market sorts women and men into different jobs with different characteristics can be expected to offer greater purchase for more specific theories when conducted in a framework independent of gender composition per se.

A Brief Review of the US Literature [3]

Our focus here, as in the majority of the literature, is upon wage studies; in particular, those investigating the impact of occupational feminization on *individual* earnings.[4] The linking theme is biases resulting from unmeasured variables,[5] but to set the scene we first review the standard approach used in an early study by Sorensen (1990), which offers a test of the crowding

[3]Our discussion offers a compact review of traditional and largely economic treatments of the effect of gender composition on wages. For newer and more nuanced explanations, see Bertrand (2011) and Blau and Kahn (2016); for studies in the sociological tradition, see in particular the work of England (e.g. England, 1982; England, Hermsen, and Cotter, 2000; England, 2005; and Levanon, England, and Allison, 2009); and for a survey of other-country studies and a comprehensive analysis of three nationally-representative British datasets – the British Household Panel Survey, the Labour Force Survey, and the Skills Survey – see Perales (2010, 2013).

[4]A useful summary of seven early studies using either unweighted or weighted occupations as the unit of analysis is provided by Sorensen (1990: Table 1). For non-wage studies of promotions within occupations and an analysis of the assignment of job points to occupations, see Paulin and Mellor (1996) and Schumann et al. (1994), respectively.

[5]In what follows, we do not consider potential selection biases. But for an excellent early study that controls for selection into employment on the part of males and females, see Blau and Beller (1988) who examine earnings differentials by gender using CPS data. The authors report that the female-male earnings differential increased over time. The selection coefficients are negative (positive) for men (women), implying that nonparticipants had higher (lower) wage offers than those in employment. In addition to gender composition, selection emerges as key to this improvement. Selection explains a large part of the improvement for white women since the increase in the selectivity of the wage regression was found to be greater for men than for women. Women also earned modestly more than men with similar characteristics in 1981 than 1971 (again selectivity-adjusted estimates)

hypothesis using data from the 1984 Panel Study of Income Dynamics (PSID) and the May/June 1983 CPS. Three different earnings models are estimated. The first is a standard human capital model augmented by the gender composition of the worker's occupation (the proportion of women in that occupation); the second adds a wider array of explanatory variables (including union status); and the third adds detailed industry dummies. The coefficient estimates for the gender composition variable decline with each augmentation but remain statistically significant throughout. For the full model, female earnings are reduced by 23 percent in the PSID – and by 15 percent in the CPS – if they are employed in an exclusively female occupation rather than an exclusively male one. The corresponding values for male earnings are decreases of 24 percent and 25 percent, respectively. And, again for the full model, the proportion of the earnings gap explained by feminization is 23 percent for the PSID and 20 percent for the CPS. These are sizable estimates (cf. Johnson and Solon, 1996), but the author cautions that if feminization is correlated with unobserved productivity characteristics it might overstate the impact of crowding.

The three remaining studies considered here return to the issue of biases in estimating the effect of feminization on earnings. One approach to the problem is that adopted by Groshen (1991), who first attempts to separate out the effect of segregation by occupation from that associated with firm and job cell (an individual's job cell being defined as all workers in the same job classification at the same establishment). Using cross section data for five industries from the BLS Industry Wage Surveys, 1974-1978, Groshen regresses the log wage on proportion female in the occupation, proportion female in the establishment, and proportion female in a job-cell, together with an individual female dummy. Multiplying the coefficient on proportion female by the gap between the proportions of female and male employment in the occupation, establishment, and job cell allows her to estimate their contribution to the observed wage gap between men and women. She reports that the largest contribution to the wage gap is the occupational component, because occupations are highly segregated and their wages are strongly linked to proportion female. Controlling for the segregation of the establishment and the job cell, the proportion female in an occupation produces a wage difference of between 11 and 26 percent, corresponding to between one-half and two-thirds of the observed wage gap. For their part, wage differences resulting from establishment and job cell segregation – even if the genders were evenly represented across occupations – sum to 12 percent, split more or less evenly between establishment and job cell (segregation). Women who work together with men in a job cell earn only one percent less, although such integration is rare. Even in integrated occupations, people work primarily with members of their own sex, and this segregation tends to raise men's wages and lower women's wages.[6]

For Groshen gender segregation accounts for most of the gender wage gap. However, very different results are reported by Bayard et al. (2003) for a much larger matched employer-employee data set. The authors seek to estimate the contribution of gender segregation by occupation, industry, establishment, and job cell to the gender wage gap, using data from the 1990 Sample Edited Data File containing worker records matched to the 1990 Standard Statistical Establishment List of establishment records (n = 637,718). Their basic regression results indicate that although the segregation of women into lower paying occupations, industries, establishments, and job cell (i.e. occupations within establishments) accounts for a material share of the gender wage gap rather more is attributable to the individual's gender. This would call for a different type of policy activism, namely equal pay legislation. Specifically, for a specification that includes a female dummy variable, four *percent-female* variables (namely, the proportions female in the occupation, industry, establishment, and job cell) as well as standard human capital and demographic controls, it is found that after accounting for segregation the gender difference in wages remains large and explains 51.4 percent of the gender wage gap. For its part the contribution of the proportion female in the job cell is 14.1 percent, while occupational segregation explains just 5 percent of the wage gap. The contributions of industry and establishment segregation are 11.23 and 15.6 percent, respectively. These results are based on highly aggregated occupations (n = 13) but raising their number to 72 still leaves a sizable gender difference in wages amounting to 40.2 percent of the wage gap. However, Bayard et al. are unable to provide longitudinal evidence that would indicate how wages change with FEM as a result of individual worker movements *between* establishments or to changes in FEM within establishments, which calls for longitudinal analysis of matched employer-employee

[6]Focusing on occupation, Groshen seeks finally to determine which of the two main theories – human capital or discrimination – is most plausible by adding measures of union status, region, general education, vocational training, strength, as well as physical demands and quality of environment for each occupation to the wage regressions previously only containing gender variables. These job attributes had little effect on the estimated coefficients for occupational gender composition.

data.[7]

A more conventional approach to tackling unmeasured variables is, then, to estimate a fixed effects model of earnings and feminization. Gerhart and El Cheikh (1991) use data from the National Longitudinal Study of Youth (NLSY) for the two years 1983 and 1986 when respondents were aged between 18 and 25 and 21 and 28 years, respectively. The authors provide both cross section and fixed effect wage estimates. The underlying earnings function includes the percentage of women in the individual's 3-digit 1970 occupation as well as a number of other occupational characteristics, individual characteristics, and industry and year dummies. Focusing on their results from the (2-year) longitudinal sample, the authors' pooled cross section or between groups model suggests that a movement from a 100 male to a 100 percent female occupation is associated with a 21.6 percent decrease in earnings for men and a 5.8 percent decrease for women. But the fixed effects or within-group estimator reduces the percentage female coefficient by a third while also rendering it statistically insignificant in the process) although the male coefficient is unchanged. The suggestion is that when fixed effects are added to models that control for occupation and industry, the impact of feminization in cross section may have more to do with (differences in) the types of people who choose to work in the more feminized occupations. Finally, when the authors decompose earnings differences into the components due to percentage female, individual characteristics, and the remaining variables (occupational characteristics, industry dummies and intercept) it is apparent that the individual and other characteristics and other variables dominate.

Despite its vintage, the final study considered here represents the most extensive

[7]An exemplary study of how much of the gender gap results from the segregation of workers across firms and jobs is Cardoso et al. (2016), who provide refined estimates of the gender wage gap filtered from the effects of job title heterogeneity and firm heterogeneity. Using a unique Portuguese dataset, these authors first estimate worker, firm, and job-title fixed effects simultaneously in wage regressions. To obtain the portion of the gender wage gap attributable to each component, they then apply Gelbach's decomposition. Essentially, differences in the fixed effects by gender capture the effect of the allocation into jobs. It is reported that the segregation of workers across firms (occupations) accounts for 18.7% (18.5%) of the conditional wage gap, meaning that men are more successful than women in sorting into firms and jobs of different quality. A subsidiary finding is that, although the two allocations have almost identical effects on the gender pay gap, the glass-ceiling effect operates mostly through worker allocation to firms, with lower access of women to higher paying firms, rather than through worker allocation to jobs; that is, not only are women sorted more frequently into lower-paying firms but the wage penalty increases with the size of the firm fixed effect. Thus, 62.8% of the gender gap persists within jobs and firms for workers of the same age and seniority, implying wage discrimination per se. See also Busch and Holst (2011) for an interesting fixed effects German study on the influence of firm size on the penalties experienced by managers in predominantly female occupations.

evaluation to date of the role of gender composition in wage determination and is perhaps most representative of the current state of play in this area of research. Macpherson and Hirsch (1995) use nationally representative national samples from the January 1983 through December 1993 monthly CPS Surveys, offering unusually large sample sizes (the total sample size is 1.84 million), in addition to various CPS supplements. The authors examine changes over time in the gender composition of jobs and its evolving effect on wages and the gender gap. The authors also estimate longitudinal wage change models for matched worker-year pairs from 1983/84 to 1992/93, now representing 25 percent of the size of the full sample.

Wage level results from the authors' standard model, containing individual characteristics, location, and broad occupation and industry, indicate that the gender composition (proportion female in the worker's 3-digit occupation) effect is large and of roughly the same absolute magnitude for both genders.[8] Expanded wage regressions containing job characteristics, such as mean years of required occupational training, computer usage, and indices of physical demands, produce much reduced gender composition coefficients – of roughly one-quarter (one-half) for women (men) – pointing to the influence of compensating differentials and/or quality sorting on the job characteristics associated with gender composition. In a final application of the wage level analysis, the authors examine the contribution of gender composition to the gender wage gap. For the standard model, gender composition accounts for more than half the explained portion of the gap, although this is reduced by about one-third for the expanded model.

Despite the importance of controlling for detailed job characteristics, there remains the issue of unmeasured skills and tastes. Here the authors' longitudinal analysis based on two-year panels of individuals seems to point to the decisive influence of person-specific labor quality and/or preference differences. Thus, for the standard model, estimates of the effects of gender composition are reduced by roughly one-half using wage change equations. For the expanded model, where the effects of feminization are already much attenuated, the gender composition coefficient estimates are now just -0.055 for women and -0.034 for men. That is to say, unmeasured skills/tastes when added to job characteristics explain some two-thirds of the standard gender composition effect among women and four-fifths of the effect among men.

[8]Suggesting that wages are 7% lower for each in a typical female occupation than in a typical male occupation, or, equivalently, that a movement toward equality of gender composition would lead to a 3.6% increase in average wages for women and a 3.4% decrease for men.

Expressed in terms of the wage gap, gender compositional differences explain just 0.02 log points of a wage gap that averaged 0.30 log points over the 1983-1993 period. In short, the gender composition variable is "correlated with differences in job characteristics, worker-specific productivity differences among observationally-equivalent workers, and taste differences regarding job characteristics" (Macpherson and Hirsch, 1995: 455). The authors thus conclude that predominantly female jobs pay less to women (and men) mostly by reason of their skill-related characteristics and quality sorting with the unmeasured skills of both genders increasing in the proportion of males in an occupation.

Because of its representativeness, use of an extensive set of variables (including occupational skills and job [dis]amenities), and complementary longitudinal analysis the Macpherson-Hirsch study provides the motivation for the present paper. We update the analysis of CPS data to determine whether the study's (cross-section) findings continue to hold. Moreover, since the most optimistic results with respect to feminization have a basis in longitudinal analysis, and given the limitations of the CPS in this regard (e.g. longitudinal data limited to two consecutive years might be inadequate for significant mobility to occur), we shall follow Gerhart and El Cheikh (1991) in using the NLSY79 to supplement the updated CPS component of our longitudinal analysis.

Econometric Specification

In our econometric modeling, and in step with Macpherson and Hirsch, we include many individual, job, and occupation-related characteristics contributing to differences in productivity and in human capital accumulation that may explain some of the wage disparity across genders. However, there is still room for unobserved taste and productivity differences, and we will need an econometric setup that accommodates such unobserved factors. Panel data methods are used to control for these unobserved time-invariant individual-specific effects.

We estimate for each gender:

$$\log\left(W_{ift}\right) = \theta_f FEM_{ift} + X'_{ift}\delta_f + Z'_{if}\beta_f + v_{ift}$$

$$\log(W_{imt}) = \theta_m FEM_{imt} + X'_{imt}\delta_m + Z'_{im}\beta_m + v_{imt} ,$$

where *f* and *m* are gender indicators; the *i* and *t* subscripts designate individuals and time, respectively; log(W) represents the natural log of hourly wages; FEM is an indicator of the proportion of women in each individual's occupation; X is a vector of observable time-varying individual-, job-, and occupation-level variables; Z is a vector of observable time-invariant characteristics; and θ, δ, and β (now excluding the gender indicators for simplicity) are the coefficients of interest. The error term v_{it} (for both genders) can be decomposed in the following way:

$$v_{it} = u_i + \varepsilon_{it} ,$$

where u_i represents individual-specific time-constant unobservable effects, and ε_{it} is a stochastic error term. In the wage equations being estimated it is very likely that unobserved factors such as ability or taste are correlated with gender composition and with other observed determinants of the wage. For this reason, our preferred model is a fixed effects specification which allows for such correlations.

Abstracting from the panel components CPS-MORG data, however, we first estimate the above models by year (and across years) and by gender, with and without human capital controls. We later add to these simple cross-sectional (yearly and pooled) models an extended set of job controls in an attempt to gauge the importance of occupation and industry level characteristics. We employ a first-differenced wage regression as a complement to OLS regression to evaluate the extent to which the relationship between occupational feminization and wages is robust to the presence of unobserved individual heterogeneity which is potentially correlated with the observed factors.

We supplement these estimates using the NLSY79. In exploiting the panel nature of this data set, in addition to first-differenced (wage change) models, we shall also estimate fixed effects models allowing correlation between unobserved individual heterogeneity and observed individual and occupational characteristics. To repeat, by virtue of its much greater number of time periods, use of the NLSY79 enables us to achieve identification through a richer source of within-group variation.

The cross sectional and longitudinal evidence on occupational feminization and wages is examined in turn. Before considering the longitudinal evidence, however, we investigate the sensitivity of the relationship between FEM and wages to model selection and detailed occupation and industry level controls. We also investigate how this relationship varies by different educational, demographic, and broad occupational subgroups. A decomposition exercise seeking to determine just how much of the wage gap is explained by the gender composition argument concludes the cross section analysis.

Data Sources and Research Sample Construction

We construct our sample from the CPS Merged Outgoing Rotation Groups (CPS-MORG) for the years 1996-2010.[9] In these data, each household is interviewed 8 times over 16 months; specifically, for an initial 4 consecutive months, followed by 8 months out of the sample, and then for a final 4 consecutive months. Beginning in 1979, households in their fourth and eighth month in the sample (i.e. the outgoing groups) were administered an earnings supplement that includes questions on their usual weekly earnings, hourly wages, usual hours worked, and (from 1983) union status on the current primary job. Approximately 60,000 households are interviewed monthly in this 4-8-4 rotation sequence.

Our CPS-MORG sample is restricted to workers aged 16 years or more. We do not consider full-time students, the self-employed,[10] or those who work for no pay. The military sample is also excluded. The wage measure is hourly wages (namely, usual weekly earnings divided by usual hours worked) that are adjusted to December 2010 dollars using the monthly Consumer Price Index for All Urban Consumers (CPI-U). As in Macpherson and Hirsch, observations with real hourly wages lower than $1 are discarded. The CPS earning data are top coded. Adjusted mean earnings above the cap were assigned on the assumption that the upper tail of the earnings distribution follows a Pareto distribution, after Hirsch and Macpherson (2011: 6). Further, all individuals with imputed wages are also excluded from our analysis.[11]

In addition to the CPS-MORG, we provide additional evidence using a long panel of

[9]The data were downloaded from *http://www.nber.org/morg/annual/* on August 11, 2014.

[10]Self-employed workers cannot be included in the analysis because they are not asked the earnings (or union) questions in the MORG earnings supplements.

[11]Given that our main interest is to extend and update Macpherson and Hirsch (1995), in previous versions of this paper we began where their analysis ended and constructed our sample from the CPS Merged Outgoing Rotation Groups (CPS-MORG) for the years 1993-2010. However, in light of (subsequent) research findings on the severity of the bias in wage effect estimates resulting from the inclusion of individuals with allocated or imputed wages in the CPS-MORG (Hirsch and Schumacher, 2004; Bollinger and Hirsch, 2006), we excluded them from our sample. As there are no valid allocation flags for imputed wages from January 1994 to August 1995, we further restricted our CPS-MORG by focusing only on the period from 1996 where we can reliably identify and exclude such observations – although we provide selected regression results for the sample that includes individuals with imputed wages in our online appendix (subsequently referenced in detail). We thank Barry Hirsch for his advice and guidance regarding this issue. All sampling weights are adjusted to rebalance the data considering the probability of non-response when we exclude imputed earners.

individuals from the core cohort of the NLSY79 for the years 1993 to 2010. The NLSY79 provides a nationally representative panel of data for the cohort of individuals aged 14 to 22 years in 1979, and who have been interviewed regularly since that year. In addition to the military, the core data exclude the oversample of Hispanic, black, and low income youth. As in the case of the CPS-MORG sample, we exclude those individuals who are self-employed or who work for no pay or who report hourly wages of less than $1. Having also excluded those with missing information on any of the variables used in the analysis, as well as observations where the wage entries are clearly in error,[12] our final sample comprises 32,957 person-year observations over the survey period analyzed. In addition to its long panel nature, use of the NLSY79 has two other advantages. One is that it allows us to track workers' actual labor market experience, which corrects for the potential measurement error in the standard experience indicator based on age and education. Another is that it also allows us to control for ability through the Armed Services Vocational Aptitude Battery (ASVAB) test scores that are unavailable in the CPS-MORG.

Although labor market activity has been surveyed in CPS – and in great detail in the NLSY79 since its inception – the occupational and industry codes are not recorded consistently across each wave of either survey. Until 2002, occupations in the CPS were recorded using the 1990 Census Occupational Classification (COC)[13] while in the case of the NLSY79 the 1980 COC was used.[14] After this year, both datasets use only the 2000 COC to code occupations. Similarly, industries are described by their 3-digit 1990 Census Industry Classification (CIC) in the CPS[15] until 2002 and by 3-digit 1980 CIC in the NLSY until 2000. Thereafter, the industries are measured by 4-digit 2002 census code in both the CPS and the NLSY79. We mapped these

[12]For example, we have a few observations for which individuals experienced wage growth of more than 100%, followed by huge declines in the very next period that were unaccompanied by any material changes in job characteristics.
[13]The variable that captures the occupation codes is designated as occ80 in the data. However, this is a misnomer. The technical appendix shows that the occupation codes were changed to COC1990 codes in 1992 (see *https:// cps.ipums.org/cps/resources/earner/cpsxNBER.pdf*, page 35).
[14]In the NLSY79 for the year 2000, occupations are measured by 1980 codes and for 2002 they are measured by 2000 census codes. In the CPS for the years 2000, 2001, and 2002 the occupations are measured by both 1990 and 2000 census codes. The mapping for this paper uses the 1990 codes for these three years.
[15]Similar to the problem with occupation coding, in the case of industry codes the variable is designated as ind80 in the data, although the CPS technical appendix shows that industry codes were changed in 1992 (*https:// cps.ipums.org /cps/resources/earner/cpsxNBER.pdf*, page 34).

occupation and industry codes so as to be able to study the full extent of the data panel available to us.[16] As in Macpherson and Hirsch, the occupations are divided into 6 separate aggregated groups and the industries into 13 one-digit groups plus the public sector.[17]

We supplemented both datasets with occupational characteristics obtained from the Occupational Information Network (O*NET) and the Occupational Projections and Training Data (OPTD) databases,[18] together with additional 3-digit industry and occupational level controls from the CPS supplements. The O*NET data provide information on strength and computer interaction requirements in each occupation, as well as occupational hazard levels and physical and environmental conditions. Besides working conditions and computer skills, we used occupational education categories from the OPTD, capturing workers' levels of human capital accumulation to include schooling and job training. The proportion of workers in large firms (i.e. those with more than 1000 employees) was calculated from the 2003-2007 CPS Annual Social and Economic (ASEC) supplements, weighted by the supplement weight; while average job tenure for each occupation was generated using the 2004-2010 CPS job tenure supplements, again weighted by the relevant supplement weight. Annual levels of union membership for each industry were obtained from *unionstats.com*. Part-time employment shares were calculated from the full CPS monthly files and weighted by the earnings weight.[19] Finally, our main control variable FEM measures the female intensity of an occupation, namely the share of female

---

[16]We use do-files kindly provided by David Macpherson to create a program to map the 2000 occupation codes into 1990 and 1980 codes. For some of the missing occupations, we updated Macpherson's crosswalk using the distribution of COC1990 to COC2000 (using *https://usa.ipums.org/usa/volii/occ_ind.shtml*). Macpherson's do-files also map and group industry classifications. For the CPS (NLSY79) the final analysis sample contains about 500 (450) unique occupations. Some 14 industry groups were generated once all the crosswalks were completed using the guidelines provided in *http://www.census.gov/people/io/methodology/*. Observe that Blau et al. (2013a, 2013b) use gender-specific crosswalks and find that gender segregation is underestimated if aggregate mappings are used. Given that the segregation indices calculated by either method are not very different for the period studied here, we eschewed the use of gender-specific crosswalks.

[17]The 6 occupations are: *managerial and professional specialty; technical, sales, and administrative support; service; farming, forestry, and fishing; precision production, craft, and repair; and operators, fabricators, and laborers*. So as to avoid identification issues in coding these industry/sectors we captured public employees via a *public administration/public sector* dummy. The remaining 13 (private) industry/sector groups are: *agriculture, forestry, and fisheries; mining; construction; manufacturing (non-durable goods); manufacturing (durable goods); transportation, communications, and other public utilities; wholesale trade; retail trade; finance, insurance, and real estate; business and repair services; personal services; entertainment and recreation services; and professional and related services.*

[18]The O*NET variables are those used in Hirsch and Schumacher (2012). The O*NET extract is from 2008 and the OPTD extract is from 2002. These datasets use COC2000.

[19]The additional CPS data were downloaded from the Integrated Public Use Microdata Series **(**IPUMS) website.

workers in the relevant 3-digit occupation. It was calculated in the same way as the part-time employment shares and from the same data source.

Cross-Sectional Evidence

In Table 1 we report sample sizes and mean wages by gender, as well as the female-to-male wage ratio and the occupational segregation index (Duncan and Duncan, 1955) for each year of the CPS-MORG data examined. As expected, the average wages of men (ranging from $21.42 in 1996 to $24.99 in 2010) exceed those of women ($16.45 in 1996 and $19.77 in 2010). The occupational segregation index is calculated as $\frac{1}{2}\sum|m_j - f_j|$, where m and f are the shares (in percent) of the male and female labor force in occupation j. Feminization (FEM) levels are reported separately by gender. The last two columns of the table give the estimates for θ_f and θ_m, namely the log wage regression coefficients on FEM without any other controls. The coefficients point to a strong unconditional negative relationship between FEM and the average female wage, and a weaker but positive association between FEM and the average male wage.

(Table 1 near here)

The relationship between FEM and wages may be capturing differences in pre-labor market and on-the-job investments to human capital, as well as reflecting the labor market attachment of the type of workers that sort into male, female, and integrated jobs and the specific characteristics of these jobs. In Table 2, therefore, we first illustrate, using a *standard* setup, how human capital, demographic and geographical differences along with broad industry and occupation contribute to the negative FEM wage relationship observed in Table 1. This initial specification controls for years of schooling, potential experience (measured by age-schooling-5) and its square, and dummies for union coverage, public sector employment, large metropolitan area, full-time employment (usual hours worked of at least 35 hours), ethnicity/race (3), marital status (2), region (8), industry/sector (13), and occupation (5). Following Macpherson and Hirsch this setup is labeled as *standard* because it uses conventional controls available in many datasets.

In an *expanded* specification, we then include an additional 8 (2) occupational (industry) attributes. The occupational attributes are (indices of) environmental (dis)amenities (*environment*), job hazards (*hazard*), strength requirements (*strength*), physical demands (*physical*), computing skills (*computers*), educational and job training requirements (*education & training*), average job tenure (*occupation tenure*), and the proportion of part-timers (*occupation* *part-time*) in each occupation. The industrial characteristics are the proportion of workers in firms having more than 1,000 employees (*industry large firms*), and the proportion of employees who are union members (*industry union*). The inclusion of these controls is intended to capture the degree to which wage differences are compensating for job (dis)amenities as well as indicating possible entry barriers impacting women's occupational choices; for example, there are more part-time jobs in predominantly female occupations in which women are also overrepresented which is suggestive of less specialized human capital accumulation in these jobs. Investments in specialized human capital (firm, industry, or job specific) are usually associated with positive wage premia.

(Table 2 near here)

FEM coefficients in the standard specification(s) are only 3 to 5 percentage points larger in absolute magnitude for women compared with the most parsimonious specifications containing only the FEM control. In the case of men, however, the effect of gender composition is now radically different in these two specifications. Moreover, feminization effects are always stronger for men than women. This result illustrates the importance of controlling for standard human capital measures and job characteristics, as well as the likely nonlinear association between FEM and wages for men (i.e. as most predominantly male jobs are low skill jobs with low wages). More importantly, it indicates that the negative relationship between gender composition and wages is substantial for both genders. From this standard regression the most important finding of all is the result that basic human capital differences do not explain the negative FEM wage relationship. That is, one cannot argue on the basis of observables that there is negative quality sorting through the FEM spectrum for either women or men: women and men working in sectors with a higher share of female workers are not necessarily lower productivity type workers earning correspondingly low wages.

When additional occupational and industry variables are controlled for in the expanded model, the FEM coefficients are reduced in absolute terms by one-half in the case of women and by about one-third for men.[20] This indicates that compensating differentials or differences in specialized human capital accumulation across occupations and industries can in fact explain a

[20]The FEM effect remains larger for men than women in the expanded model, just as it was in the standard model. This is different from Macpherson and Hirsch's findings. In their study period even though estimates from the standard model are mostly higher for men (their Table 3) this pattern is reversed in the expanded model.

significant portion of the negative relationship between gender composition and wages in an occupation. Comparing these results with those obtained by Macpherson and Hirsch (their Table 3) reveals generally larger negative FEM coefficients for both genders (quantitatively compared below). However, as they also note, even though the market has become less segregated over time, this does not translate into a reduction in the wage gap as the wage penalty associated with higher female presence in a job has also increased. We might therefore conclude that younger cohorts experience higher wage penalties in female jobs. This outcome may not necessarily be indicative of discrimination but instead reflect unobserved characteristics of these jobs or the individuals concerned. Both aspects are further considered in our longitudinal analysis.

By way of summary, according to the estimates for the standard model, average wages are 9.9 percent [(0.666-0.313)*(-0.281)] lower for women and 11.6 percent [(0.666-0.313)*(-0.328)] lower for men in typical female jobs compared with typical male jobs in 2010. Expressed differently, a non-segregated market – with a 0.49 female presence in each job, which is the mean value of FEM in the combined sample as of 2010 – would improve female wages by about 5 percent on average while lowering male wages by somewhat more.[21] Corresponding measures from the expanded model imply a 2.7 percent increase for women and a 4 percent decrease for men.[22] That is, a significant portion – but not all – of the FEM effect can be explained by occupational and industry level differences across jobs. We will subsequently investigate which specific components are the main drivers of the gender composition effects.[23] Before proceeding with such an analysis, however, we will first investigate the linearity of the FEM wage relationship. Because the most extreme male jobs are low skill and poorly paying, the unconditional means for men do not capture the negative effect of FEM on the rest of the wage distribution. We will next check if such non-linearity persists in the standard and expanded models.

(Table 3 near here)

---

[21]Specifically, by (0.49-0.666)*(-0.281) and (0.49-0.313)*(-0.328), respectively.

[22]In the text we only report results for a sample that excludes imputed earners. In our online appendix (https://sites.google.com/site/orguldemetozturk/research2/the-occupational-feminization-of-wages), we present coefficient estimates for the sample including individuals with imputed wages. See the spreadsheets *Table 2 Full Sample* through *Table 4 Full Sample*. As predicted by one of our referees, inclusion of imputed earners creates substantial bias in the estimates; specifically, the FEM coefficient estimates are significantly lower in absolute value when these individuals are included in the estimation sample.

[23]Coefficient estimates for these specific components of the standard and expanded models of Table 2 are again provided in our online appendix (see the spreadsheet *Complete Results for Table 2*).

In Table 3, we pool the data and treat feminization in the first instance as a continuous variable (see ‘Model 1’). This definition of FEM is the same as that in Table 2, albeit now estimated with pooled data rather than annual data. We then specify feminization as a categorical variable (‘Model 2’), with dummies for each quartile. The omitted category comprises predominantly male jobs. This estimation strategy accommodates two facts: (a) that the distribution of feminization among the two groups sharply differs (for example, average FEM across all years among women is 0.666 whereas it is only 0.302 for men); and (b) that both predominantly male and predominantly female jobs are low paying. We might therefore expect the relationship between feminization and wages to be nonlinear for both genders. However, as is evident, although the relationship is somewhat inverse U-shaped for the no-controls specifications, with richer models the relationship becomes more or less linear. In all models, moreover, the most severe wage penalties are associated with the highest levels of feminization for each gender. Even though in the expanded model female workers in the second FEM quartile earn significantly higher wages than all others, the difference with respect to the first FEM quartile is relatively small in comparison. To this extent, our results again correspond to those of Macpherson and Hirsch (their Table 4). Accordingly, we will continue to use a linear specification in our analysis.

(Table 4 near here)

In order to establish which characteristics contribute most to the FEM-wage relationship, we next examine in Table 4 the sensitivity of the gender composition argument across somewhat more differentiated specifications. Observe firstly that the addition of broad industry categories to a base set of individual characteristics (comparing line 3 with line 2) little affects the FEM coefficient estimate for women but materially reduces it in the case of men (from -0.176 to -0.103), at the same time as the inclusion of occupational dummies strongly increases that negative coefficient to -0.367 (line 4). These results for men indicate that much of the inverse association between gender composition and wages is accounted for by industry differences and occurs primarily within broad industry groups. This is not the case for women.

Taking the standard model (line 5) as the comparator, it can be seen that for both genders, but especially women, occupational education and training requirements (line 6) and the share of part-timers in an occupation (line 8) are most influential. For their part other occupational characteristics (occupational tenure, computer use and physical [dis]amenities) seemingly explain very little. In sum, these results indicate that once we control for industry and occupational categories a sizable portion of the negative FEM-wage relationship is due to occupational differences in skill requirements and job attachment (comparing lines 10 to 5). This is true not only for women but also for men. Although controlling for industry proportion unionized and the share of large firms reduces the absolute value of the FEM coefficient for women, the increase reported for men is barely discernible (comparing lines 12 and 5). Taken in the round – that is, now referring to the expanded model in line 13 – controlling for these occupational and industry characteristics reduces the negative FEM coefficient estimate by 52.7 percent for women and 37.5 percent for men. Observe that very little to none of this effect is produced by the physical requirements of the occupations (line 14).

(Table 5 near here)

Our standard and expanded models are estimated across different educational, demographic, and occupational groups in Table 5. This exercise enables us to understand differential effects of occupational gender composition on wages and helps shape policy design. Beginning with age, the most negative effects are found for 30 to 39 year olds closely followed by those aged 40 to 49 years. This pattern is true for both women and men. The overall relationship between age and gender composition is somewhat U-shaped. Individuals may be sorting into female jobs when they need flexibility for fertility reasons or to care for elderly relatives, responsibilities that also reduce productivity. Timing of these events very likely overlaps with the mid-career years, when occupational investments such as training or longer working hours may yield the highest wage returns. With respect to marital status, for both women and men the most negative effect is observed when the individual is married with a spouse present. This result might be mirroring the relationship earlier observed for the age groups. In the case of education, women with the highest education levels (16 or more years of schooling) are the most damaged by gender composition of the job. Since they are more likely to have access to well-paid mixed gender and mostly male occupations, individuals in this group may be trading large wage returns for greater flexibility and more family friendly aspects of the jobs. The negative effects are greatest among men with a college degree, although for those men with higher than a baccalaureate degree the effect is muted in the expanded model and is only about one-quarter of what it was in the standard specification. For this group of men, FEM seemingly captures job amenities and/or a lack of specialized human capital more so than it does for their female counterparts.

As far as race is concerned, the adverse effect of feminization is lowest among (Non-Hispanic) blacks and Hispanics of both genders, although this outcome may of course be capturing the lack of opportunities available to these minorities in high wage markets. The negative gender composition effect is larger in the union sector in the standard specification. However, in the expanded models this pattern disappears, and for women in union jobs there remains no significant FEM effect. The FEM coefficient is more negative for both women and men in the public sector across specifications. The gender composition penalty also applies generally to full-time work, and is larger for men than women. In fact, the FEM coefficient is positive but not significant for women in part-time jobs in the expanded model, although it remains modestly negative for men.

Finally, at the base of Table 5 are given the FEM coefficients by broad occupational groups. The first three occupational groups contain large numbers of both genders, facilitating comparisons between them.[24] Here, FEM effects are most negative for men and women in *technical, sales, and administrative support*, but for both genders the coefficients are much reduced once we control for occupational and industry-level characteristics in the expanded model. Even though the negative effects of FEM are smaller for both genders in the *managerial and professional specialty* occupations, they are relatively more persistent for women in the expanded model. This persistence may be reflecting structural differences and/or a high degree of segregation within this occupational grouping. Among managers, shorter authority ladders in female jobs may account for this persistence, while in the case of professional specialists there are highly segregated and differentially paid occupations (e.g. economists/social workers and nurses/physicians).

(Table 6 near here)

We earlier examined the sensitivity of the FEM coefficient to choice of model structure and component. In Table 6 we decompose the log wage gap between men and women, now

[24]The remaining occupational groups are insufficiently representative to make meaningful comparisons, but once we draw a distinction between production and non-production occupations our results mirror those of Macpherson and Hirsch.

exploring the sensitivity of the gender wage gap by specification and year to the inclusion of FEM.[25] Even though a job's gender composition explains more of the wage gap than all other job characteristics combined, the main impression conveyed by Table 6 is the very scale of the unexplained part of the gender wage gap. That is, even with a very full set of human capital and job controls, some 60 to 70 percent of the wage gap remains unexplained. Over time it is also evident that the importance of observed job attributes in explaining the wage gap declines. This outcome may result from an increased importance of unobserved individual differences in productivity/tastes or it may reflect discrimination or unobserved job attributes that are correlated with FEM. In next modeling unobserved individual heterogeneity using longitudinal data, we will directly explore the relevance of unobserved productivity explanation, returning to the role of discrimination and unobserved job attributes and tastes in our concluding remarks.

Longitudinal Evidence

In this section we probe the role of unobserved factors by exploiting the panel nature of the CPS-MORG. In addition, we run a wider set of panel models using our NLSY79 sample, the main contribution of this paper being not only to update Macpherson and Hirsch but also to extend their longitudinal analysis using a much longer panel that surveys individual work histories more thoroughly.

Beginning with the CPS, we have first to discuss the construction of the panel and the issues of measurement error and generalizability. We use a similar method to Macpherson and Hirsch in constructing our panel. Individuals in the CPS-MORG can in principle be matched across two consecutive years. To be considered a valid match, the individual in the 8th rotation group should have an identical household identifier, survey month, line number, and state to an individual in the 4th rotation group. As the survey can take place on different days in the same survey month, the age difference does not have to equal one and so we allow for an age difference of between zero and two years. However, the matched pair should have the same individual characteristics such as race and gender. Following Macpherson and Hirsch, we also deleted "bad" matches recording incorrect changes in marital status (from married to never

[25]The (twofold) decompositions in Table 6 are performed using the Stata *oaxaca* command with the pooled option and a group indicator, as recommended by Jann (2009). Alternative decompositions in which either men or women are used as the reference group or where the weights are the proportion of each gender in the sample (in the manner of Macpherson and Hirsch) are presented in our online appendix (see the spreadsheet *Table 6 Alternatives*).

married), ethnic status (a change in Hispanic status), education (a change in schooling other than zero or one years), and in veteran status. Because the CPS surveys identify residences not households, we cannot match those individuals who do not occupy the same residence in both years. The match yields 398,745 observations from 201,655 women and 197,090 men prior to occupational and industry change.

Next, even in the absence of individuals with allocated earnings in the CPS-MORG sample, measurement error arises because a sizable share of recorded occupation changes are not true changes but instead the result of Census coders in successive years assigning different occupations based on workers' varying descriptions of their jobs. In these circumstances, restricting the CPS-MORG sample to those who change occupation and industry may be expected to materially reduce measurement error in the change in occupational FEM. That is to say, workers who report changes in both industry and occupation are less likely to stay in the same job than those workers formally recorded as changing (only) their occupation.[26] Finally, by construction the CPS-MORG panel is going to exclude individuals who are changing households, who are disproportionally younger, and may have a very different occupation, industry, and FEM profile.[27] This may raise some concerns about generalizability of the panel data results.

In Table 7 we first report the pooled OLS wage level estimates in order to establish the generalizability of our results. Comparing the pooled OLS results in Table 7 with the Model 1 results from Table 3, we can conclude panel wage level estimates are indeed generalizable albeit stronger, especially for women. Compared with the cross section sample, this group has lower unionization, as well as smaller shares of public employees, workers employed in industries with large firms, and occupations with part-time employees. All such factors are associated with more negative FEM effects, as was reported in Table 5.

(Table 7 near here)

We next look at the models of wage change. Were we to obtain in Table 7 results from the wage change (FD) equations that mirror those for the OLS equations, we would conclude that

---

[26]We do not need to impose this occupation-and-industry change restriction for the NLSY79 as the occupation changes are recorded more accurately in that dataset. Moreover, most of the job changes occur in the early career years and our NLSY79 sample (being aged 31 to 39 years in 1996 and 45 to 53 years in 2010) is on this account less mobile.

[27]See the online appendix spreadsheet *Comparisons CPS-MORG Samples*.

unobserved heterogeneity cannot contribute anything beyond the explanation offered by measured human capital characteristics and job attributes, with the FEM coefficients likely capturing discrimination and resulting crowding effects on wages. But this is not the case. In the panel data models we see dramatically lower coefficients for the ΔFEM variable, especially in the expanded specification where statistical significance is also much reduced where present. These results indicate that much of the negative correlation between wages and the gender composition of a job can be explained by unobserved factors.[28] The suggestion is, then, that preference/taste differences, and possibly unobserved worker productivity, play a pivotal role in explaining away the effects of gender composition on wages.

It will be recalled that our panel data estimates are provided for occupation-and-industry changers alone to tackle the problem of measurement error. In estimates not reported here for the sample of occupation-only switchers we obtained coefficient estimates that were lower in absolute magnitude than those reported for the occupation-and-industry switchers in both the standard and expanded models (and especially the former) for both men and women. That being said, in each case the differences are modest and considerably smaller than those reported by Macpherson and Hirsch, indicating that measurement error is less of an issue for our sample.[29]

The lower portion of Table 7 presents results for the NLSY79 sample. As for the CPS-MORG, pooled OLS wage level and panel wage change gender composition estimates are provided. Moreover, the table also gives corresponding estimates from panel fixed effects models.[30] We should note at the outset that the NLSY79 panel closely resembles its CPS-MORG counterpart in descriptive statistics and with respect to the estimates derived from the cross-sectional models.[31]

Table 7 also provides results for an "expanded plus" specification in which the measure of potential experience is replaced by actual labor market experience and controls for age and *innate ability* (via age- and education-adjusted ASVAB scores) are added. Men and women of

[28]See the online appendix spreadsheet *Diagnostic*s.
[29]See the online appendix spreadsheet *Table 7 Other Panel Samples*.
[30]We also estimated random effects (RE) models. The RE model coefficient estimates are significant for men and women across the board, but are not separately provided here as the Hausman test statistics indicate that only the fixed effects model estimates are consistent. Again see the spreadsheet *Diagnostics* in our online appendix.
[31]The NLSY79 counterparts of Tables 1 and 2 for CPS-MORG are contained in the online appendix spreadsheets *Table 1 NLSY79 Version* and *Table 2 NLSY79 Version*. Basic descriptive statistics for the NLSY79 panel are given in the spreadsheet *Descriptives NLSY79*.

the same age and education may differ significantly in their labor market experience as the labor market participation of women is frequently interrupted. As a result, *potential experience* may be a poor measure of labor market attachment for women. For its part, the inclusion of age- and education-adjusted ability scores might help in distinguishing between unobserved taste and unobserved ability explanations. Unobserved ability and unobserved tastes are likely correlated with FEM in different ways, offsetting each other's bias. Does the inclusion of an actual measure of ability provide an indication of the extent of omitted ability bias in the FEM coefficient? We would expect unobserved ability to be positively correlated with wages and negatively correlated with FEM, leading to a smaller FEM coefficient estimate as a result. However, the difference between the OLS FEM coefficients is statistically insignificant for both genders, undercutting support for an unobserved ability explanation.

A direct comparison between the NLSY79 and CPS-MORG samples is of course only possible for the *first differenced* models. In comparing the FD results for the former with those for occupation-and-industry changers in the CPS-MORG, the ΔFEM coefficients are between 50 percent (in the expanded specification) and 100 percent (standard specification) higher in absolute magnitude than the corresponding estimates for the CPS-MORG samples. Similar results are obtained from our preferred fixed effects models. The stronger ΔFEM coefficients may be due to the fact that the NLSY79 is an aging cohort and reflect the heritage effects of past discrimination. This possibility will be further addressed in the next section.

Overall, panel model estimates from NLSY79 also illustrate the role of unobserved tastes or unobserved job characteristics that are negatively correlated with wages but positively correlated with FEM, as it will be recalled that we ruled out the unobserved ability explanation in the OLS models. In the expanded specifications of the fixed effects model, the FEM coefficient estimates are no longer statistically significant for men. That is to say, factoring out unobserved individual heterogeneity negates the previously statistically significant negative FEM coefficients reported for men in their mid- and late-career years. (And, to repeat, in this comparison unobserved ability does not seem to be a significant factor.) However, for the NLSY79 cohort of women, even when unobserved tastes and abilities are controlled for, there remains a log point difference of about 0.0169 in favor of typically male jobs.[32] As a practical

[32]Calculated by multiplying the FEM coefficient from the expanded FE model by the difference in feminization

matter, this represents a little over 6 percent of the difference in average log wages between men and women (of 0.262 log points in 2010).

Robustness

One might argue that the NLSY79 and CPS-MORG panels are very different in nature not only because of the length of time over which each individual is observed but also because the NLSY79 is an aging cohort with individuals who are in their mid-to-late careers during our sample period. To address this issue we create synthetic *aging* cohorts from the CPS-MORG sample. Given that the NLSY79 surveys individuals aged 14 to 22 years in 1979, we first construct a cohort starting with those individuals aged 31 to 39 years in the 1996 CPS-MORG survey. We next select individuals aged 32 to 40 years in the following year's CPS-MORG survey. In this manner, the 'aging' sample builds with a one-year increase in the age boundaries for each successive year of the CPS-MORG survey, thus mimicking the NLSY79 sample. Similarly, we constructed two additional *aging* cohorts, one ten years younger and the other ten years older than the NLSY79 cohort at its inception. Specifically, the younger aging cohort starts with 21 to 29 year-olds in the 1996 CPS-MORG and the older aging cohort with 41 to 49 year-olds, the age boundaries again increasing by one year with each consecutive round of the survey.

(Table 8 near here)

From Table 8 it can be seen that the negative effects of FEM are largest for the youngest aging female cohort. The gender composition effect for this group is on average 2.5 (4.6) times more negative than it is for the oldest aging female cohort in the standard (expanded) model. This result seemingly counters the heritage effect argument mentioned in the previous section. When we also ran these models with younger (but not aging) cohorts from the CPS-MORG panel – by restricting observations to either 20 to 30 or 20 to 40 years olds – we again obtained similarly higher FEM coefficients relative to the whole CPS sample.[33] However, these coefficient estimates are still not as strong as those for the NLSY79 cohort. We would interpret these robustness checks as indicating that discrimination while undoubtedly a factor is by no means the sole explanation. Life cycle considerations, possibly operating through unobserved job attributes, evidently also play a role.

---

rates between typically female and male jobs in 2010; that is, 0.045*(0.654-0.278) = 0.0169.

[33]Results of this exercise are given in our online appendix in the spreadsheet *Table 7 CPS 20 to 40 Year Olds.*

Conclusions

Not only are women being encouraged to enter male dominated and highly paid occupations (e.g. in *STEM* fields) but also technological advances now make it possible for them to perform many physical jobs without the exertion of physical power, thereby eliminating a previously supposed male advantage. Despite this progress, college majors differ among the genders, while women have materially higher part-time employment and shorter working weeks even when employed full-time. The seeming counterpart is that, even after many decades of increasing female presence in the labor market and evolving gender roles, we still observe male jobs and female jobs. Moreover, female jobs persistently have lower wages. Do these jobs require lower education, less experience, and less overall human capital for the reasons hinted at earlier? Or are they crowded with an excess supply of female labor that is discriminated against or excluded everywhere else, and/or do they offer a different portfolio of benefits? In the present exercise, we have examined the extent to which a higher share of women in a job contributes to observed wage differences. In seeking to understand the role of feminization, we have explored explanations such as quality sorting, discrimination, and unobserved differences in abilities and preferences.

Our results, in common with those of Macpherson and Hirsch, indicate that only a portion of the wages of men and women are explained by gender composition. The specifics are as follows. In cross section, the FEM coefficients remain significant and negative for both genders, although in the presence of the human capital and occupational controls they are reduced significantly for females. For men, on the other hand, gender composition effects for the pooled sample become more negative in the presence of these controls. The CPS-MORG panel model estimates for FEM in the case of women are small but still statistically significant, while for men they are no longer significant when we control for detailed occupational and industry characteristics. The suggestion is that both women and men tend to sort into predominantly female jobs either because of their lower unobserved skills or because of their unobserved taste differences that are correlated with gender composition and measured and unmeasured job characteristics.

Our second data set, the NLSY79, provides us with a longer panel for an aging cohort who were well into their careers at the beginning of the study period and most of whom were approaching retirement age by its end. Our estimates using this data set tell a similar story to those obtained from the CPS-MORG for both pooled and first differenced models and are confirmed with supplementary fixed effects estimates confirm these results.

Panel estimates from the NLSY79 are stronger than their CPS-MORG counterparts. One might think that this result is due to the fact that the NLSY79 is an aging cohort; that is, the magnitude of the FEM coefficients would be smaller for younger female cohorts were discrimination the dominant (unobserved) factor. However, when we constructed synthetic aging cohorts from the CPS-MORG data, we found that the effect was much stronger for our youngest cohort. It follows that further research on this discrimination component is required.

We would conclude along with Macpherson and Hirsch that policies to increase the 'female component' in male jobs through quotas and the like will not be enough to eliminate wage discrepancies. This is because the wage penalties paid by women for working in female jobs are in part compensation for non-wage job attributes such as flexible schedules or other family friendly policies. In other words, even if they had the skills to be employed in the male jobs, women may choose not to enter them on the grounds that such jobs provide insufficient flexibility, inter al. In order to increase female presence in male jobs, policies need to be directed toward addressing the (dis)amenities of male and mixed gender jobs through such measures as paid parental leave and family sick leave. As it stands, women in these jobs may be having to sacrifice more financially or are expected to accept less when seeking similar levels of flexibility and benefits that female jobs are possibly offering.

Appendix: Examples of Female/Male Occupations and Content of OPTD and O*NET Variables

(i) Extreme Occupations

Examples of female jobs (FEM given in parentheses):

> Kindergarten and earlier school teachers (98%);
> Dental hygienists (98%);
> Dental assistants (97%);
> Secretaries (97%);
> Child care workers (94%); and,
> Licensed practical nurses (94%).

Examples of male jobs:

> Heavy equipment and farm equipment mechanics (1%);
> Drillers of oil wells (1%);
> Elevator installers and repairers (1%);
> Bus, truck, and stationary engine mechanics (1%);
> Plasterers (1%); and,
> Concrete and cement workers (1%).

(ii) The OPTD Education & Training Categories

1 First professional degree;
2 Doctor's degree;
3 Master's degree;
4 Degree plus work experience;
5 Bachelor's degree;
6 Associate's degree;
7 Postsecondary vocational award;
8 Work experience in a related occupation;
9 Long-term on-the-job training;
10 Moderate-term on-the-job training; and,
11 Short-term on-the-job training.

*(iii) Content of O*NET Working Conditions Indices*

38 of the 259 O*NET variables used by Hirsch and Schumacher (2012) and Hirsch and Manzella (2015) are as follows:

Static strength	strength
Explosive strength	strength
Dynamic strength	strength
Trunk strength	strength
Stamina	strength
Frequency of conflict situations	environment
Deal with unpleasant or angry people	environment
Deal with physically aggressive people	environment
Indoors, environmentally controlled	environment
Indoors, not environmentally controlled	environment
Outdoors, exposed to weather	environment
Outdoors, under cover	environment
In an open vehicle or equipment	environment
In an enclosed vehicle or equipment	environment
Physical proximity	environment
Sounds, noise levels are distracting or uncomfortable	environment
Very hot or cold temperatures	environment
Extremely bright or inadequate lighting	environment
Cramped work space, awkward positions	environment
Exposed to contaminants	hazard
Exposed to whole body vibration	hazard
Exposed to radiation	hazard
Exposed to disease or infections	hazard
Exposed to high places	hazard
Exposed to hazardous conditions	hazard
Exposed to hazardous equipment	hazard
Exposed to minor burns, cuts, bites, or stings	hazard
Wear common protective or safety equipment such as safety shoes, glasses, gloves, hearing protection, hard hats, or life jackets	hazard
Wear specialized protective or safety equipment such as breathing apparatus, safety harness, full protection suits, or radiation protection	hazard
Spend time sitting	physical
Spend time standing	physical
Spend time climbing ladders, scaffolds, or poles	physical
Spend time walking and running	physical
Spend time kneeling, crouching, stooping, or crawling	physical
Spend time keeping or regaining balance	physical
Spend time using hands to handle, control, or feel objects, tools, or controls	physical
Spend time bending or twisting the body	physical
Spend time making repetitive motions	physical

References

Bayard, Kimberly, Judith Hellerstein, David Neumark, and Kenneth Troske. 2003. New evidence on sex segregation and sex differences in wages from matched employee-employer data. *Journal of Labor Economics* 21(4): 887-922.

Becker, Gary S. 1985. Human capital, effort, and the sexual division of labor. *Journal of Labor Economics* 3(1): 33-58.

Bergmann, Barbara R. 1974. Occupational segregation, wages and profits when employers segregate by race and sex. *Eastern Economic Journal* 1(2): 103-10.

Bertrand, Marianne. 2011. New perspectives on gender. In Orley C. Ashenfelter and David Card (Eds.), *Handbook of Labor Economics*, Volume 4b, pp. 1545-92. Amsterdam: Elsevier.

Blau, Francine D., and Andrea H. Beller. 1988. Trends in earnings differentials by gender 1971-1981. *Industrial and Labor Relations Review* 41(4): 513-29.

Blau, Francine D., Peter Brummund, and Albert Yung-Hsu Liu. 2013a. Trends in occupational segregation by gender 1970-2009: Adjusting for the impact of changes in the occupational coding system. *Demography* 50(2): 471-92.

Blau, Francine D., Peter Brummund, and Albert Yung-Hsu Liu. 2013b. Erratum to: Trends in occupational segregation by gender 1970–2009: Adjusting for the impact of changes in the occupational coding system. *Demography* 50(2): 493-94.

Blau, Francine, and Lawrence M. Kahn. 2016. The gender wage gap: Extent, trends, and explanations. IZA Discussion Paper No. 9656. Bonn: Institute for the Study of Labor. (Forthcoming in *Journal of Economic Literature*.)

Bollinger, Christopher R., and Barry T. Hirsch. 2006. Match bias from earnings imputation in the current population survey: The case of imperfect matching. *Journal of Labor Economics*, 24 (3): 483-519

Busch, Anne, and Elke Holst. 2012. Occupational sex segregation and management-level wages in Germany: What role does firm size play? IZA Discussion Paper No. 6568. Bonn: Institute for the Study of Labor.

Cardoso, Ana Rute, Paulo Guimarães, and Pedro Portugal. 2016. What drives the gender wage gap? A look at the role of firm and job title heterogeneity. *Oxford Economic Papers Advance Access*, January 7. doi: 10. 1093/oep/gpv069.

Clausen, John A. 1968. *Socialization and Society*. Boston, MA: Little Brown and Company.

Duncan, Otis D., and Beverley Duncan. 1955. A methodological analysis of segregation indices. *American Sociological Review* 20(2): 201-17.

England, Paula, George Farkas, Barbara S. Kilbourne, and Thomas Dou. 1988. Explaining occupational sex segregation and wages: Findings from a model with fixed effects. *American Sociological Review* 53(4): 544-58.

England, Paula. 1982. The failure of human capital theory to explain occupational sex segregation. *Journal of Human Resources* 17(3): 358–70.

England, Paula. 2005. Gender inequality in labor markets: The role of motherhood and segregation. *Social Politics* 12(2): 264-68.

England, Paula, Joan M. Hermsen, and David A. Cotter. 2000. The devaluation of women's work: A comment on Tam. *American Journal of Sociology* 105(6): 1741-51.

England, Paula, Melissa S. Herbert, Barbara S. Kilbourne, Lori L. Reid, and Lori McCreary Megdal. 1994. The gendered valuation of occupations and skills: Earnings in the 1980 census occupations. *Social Forces* 73(1): 65-100.

Flood, Sarah, Miriam King, Steven Ruggles, and J. Robert Warren. (2015) *Integrated Public Use Microdata Series, Current Population Survey: Version 4.0* [Machine-readable database]. Minneapolis, MN: University of Minnesota.

Gelbach, Jonah B. 2016. When do covariates matter? And which ones, and how much? *Journal of Labor Economics* 34(2): 509-43.

Gerhart, Barry, and Nabil El Cheikh. 1991. Earnings and percentage female: A longitudinal study. *Industrial Relations* 3(1): 62-78.

Groshen, Erica L. 1991. The structure of the female/male differential: Is it who you are, what you do, or where you work? *Journal of Human Resources* 26(3): 457-72.

Hakim, Catherine. 2000. *Work-Lifestyle Choices in the 21st Century: Preference Theory*. Oxford, England: Oxford University Press.

Hirsch, Barry T. and David A. Macpherson. 2011. *Union Membership and Earnings Data Book: Compilations from the Current Population Survey*. Washington, DC: Bureau of National Affairs. Available at *http://unionstats.gsu.edu/Pareto Top-Code Earnings.xls*

Hirsch, Barry T., and Edward J. Schumacher. 2004. Match bias in wage gap estimates due to earnings imputation. *Journal of Labor Economics* 22(3): 689-722

Hirsch, Barry T., and Edward J. Schumacher. 2012. Underpaid or overpaid: Wage analysis for nurses using job and worker attributes. *Southern Economic Journal* 78(4): 1096-19.

Hirsch, Barry T., and Julia Manzella. 2015. Who cares – and does it matter? Measuring the wage penalty for caring work. *Research in Labor Economics* 40: 213-75.

Jann, Ben. 2008. The Blinder-Oaxaca decomposition for linear regression models. *The Stata Journal* 8(4): 453-79.

Johnson, George, and Gary Solon. 1986. Estimates of the direct effects of comparable worth policy. *American Economic Review* 76(5): 1117-25.

Kilbourne, Barbara S., Paula England, George Farkas, Kurt Beron, and Dorothea Weir. 1994. Returns to skills, compensating differentials, and gender bias: Effects of occupational characteristics on the wages of white men and women. *American Journal of Sociology* 100(3): 689-719.

Levanon, Asaf, Paula England, and Paul Allison. 2009. Occupational feminization and pay: Assessing causal dynamics using 1950-2000 U.S. Census Data. *Social Forces* 88(20): 865-81.

Lewis, Gregory B. 1996. Gender integration of occupations in the federal civil service: Extent and effects on male-female earnings. *Industrial and Labor Relations Review* 49(3): 472-83.

Macpherson, David A., and Barry T. Hirsch. 1995. Wages and gender composition: Why do women's jobs pay less? *Journal of Labor Economics* 13(3): 426-71.

McArthur, Leslie. 1985. Social judgment biases in comparable worth analysis. In Heidi I. Hartmann (Ed.), *Comparable Worth: New Directions for Research*, pp. 53-70. Washington, DC: National Academy Press.

Magnusson, Charlotta. 2009. Gender, occupational prestige, and wages: A test of devaluation theory. *European Sociological Review* 25(1): 87-101.

Marini, Margaret M., and Mary C. Brinton. 1984. Sex typing in occupational specialization. In Barbara F. Reskin (Ed.), *Sex Segregation in the Workplace*, pp. 192-232. Washington, DC. : National Academy Press.

Meyer, Peter B., and Anastasiya M. Osborne. 2005. Proposed category system for 1960-2000 census occupations. BLS Working Paper No. 383, Washington, D.C.: U.S. Bureau of Labor Statistics.

Osterman, Paul. 1987. Choice of employment systems in internal labor markets. *Industrial Relations* 26(1): 46-67.

Paulin, Elizabeth A., and Jennifer M. Mellor. 1996. Gender, race, and promotion within a private-sector firm. *Industrial Relations* 35(2): 276-95.

Perales, Francisco. 2010. Occupational feminization, specialized human capital and wages: Evidence from Britain. Working Paper No. 2010-31, Institute for Social and Economic Research, University of Essex. Available at: *https://www.iser.essex.ac.uk/files/ iser_working_papers/2010-31.pdf*

Perales, Francisco. 2013. Occupational feminization, specialized human capital and wages: Evidence from Britain. *Work, Employment and Society* 27(4): 600-20.

Pergamit, Michael R., and Jonathan R. Veum. 1999. What is a promotion? *Industrial & Labor Relations Review* 52(4): 581-601.

Polachek, Solomon W. 1981. Occupational self-selection: A human capital approach to sex differences in occupational structure. *Review of Economics and Statistics* 63(1): 60-69.

Polachek, Solomon W. 1985. Human capital: a defense of human capital predictions. *Journal of Human Resources* 20(3): 437-40.

Ruggles, Steven J., Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. 2010. *Integrated Public Use Microdata Series: Version 5.0* [Machine-readable database]. Minneapolis, MN: University of Minnesota.

Schumann, Paul L., Dennis A. Ahlburg, and Christine Brown Mahoney. 1994. The effects of human capital and job characteristics on pay. *Journal of Human Resources* 29(2): 481-503.

Sorensen, Elaine. 1990. The crowding hypothesis and comparable worth. *Journal of Human Resources* 25(1): 55-89.

Tam, Tony. 1997. Sex segregation and occupational gender inequality in the United States: devaluation or specialized training. *American Journal of Sociology* 102(6): 1652-92.

Treiman, Donald J., and Heidi I. Hartmann (eds.) 1981. *Women, Work, and Wages: Equal Pay for Jobs of Equal Value*. Washington, DC: National Academy Press.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press.

Table 1 . Mean Wages, the Wage Ratio, Gender Composition, and the Wage-Gender Composition Relationship by Year, 1996-2010

	Women			Men			Female-to-Male	Segregation		
Year	*N*	Wage	FEM	*N*	Wage	FEM	Wage Ratio	Index	θ_f	θ_m
1996	52,040	16.45	0.670	52,895	21.42	0.295	0.768	0.531	-0.234**	-0.005
1997	53,010	16.70	0.667	53,692	21.88	0.296	0.763	0.527	-0.227**	-0.003
1998	52,468	17.26	0.666	53,477	22.21	0.297	0.777	0.527	-0.247**	0.016
1999	50,649	17.58	0.662	51,455	22.97	0.300	0.765	0.520	-0.249**	0.017
2000	49,630	17.80	0.661	50,597	23.12	0.301	0.770	0.518	-0.283**	0.043**
2001	51,867	18.33	0.660	52,839	23.83	0.302	0.769	0.516	-0.267**	0.045**
2002	56,549	18.66	0.661	56,958	24.00	0.304	0.777	0.518	-0.232**	0.072**
2003	54,042	18.85	0.668	54,037	24.06	0.303	0.784	0.524	-0.217**	0.036**
2004	53,120	18.91	0.668	53,481	24.02	0.298	0.787	0.527	-0.240**	0.067**
2005	53,745	18.92	0.667	54,335	24.03	0.300	0.787	0.527	-0.228**	0.060**
2006	53,463	18.92	0.667	54,415	23.85	0.298	0.793	0.525	-0.230**	0.072**
2007	53,868	19.18	0.666	54,768	24.15	0.301	0.794	0.523	-0.244**	0.079**
2008	53,742	19.23	0.668	54,081	24.54	0.305	0.784	0.523	-0.245**	0.069**
2009	52,980	19.68	0.669	52,293	25.15	0.311	0.782	0.518	-0.223**	0.024+
2010	50,219	19.77	0.666	49,650	24.99	0.313	0.791	0.514	-0.220**	0.026*

Notes: Data are from the 1996-2010 annual CPS-MORG files. Wages are calculated as usual weekly earnings divided by usual weekly hours and are converted to 2010 December dollars using the monthly CPI-U. The segregation index (Duncan and Duncan, 1955) is calculated by $1/2\sum|m_j-f_j|$, where m_j and f_j are the proportions of male and female employment in occupation j. FEM measures the proportion of females to total employees in the worker's detailed occupation. θ_f and θ_m are the gender-composition coefficients from the regression of log wages on feminization (FEM) with no other controls. **,*,+ denote statistical significance at the 0.01, 0.05, and 0.10 levels, respectively. Sample includes only those with (non-imputed) wage information. However, all occupation and industry level characteristics, including the segregation index, are calculated using all those with valid occupation and industry information. All regressions use CPS-MORG sampling weights which are adjusted using the inverse non-response probability whenever necessary.

Table 2. Gender Composition Coefficients from Unconditional, Standard, and Expanded Wage Level Equations, by Gender and Year, 1996-2010

	Women				Men			
Year	*N*	No Controls	Standard	Expanded	*N*	No Controls	Standard	Expanded
1996	52,040	-0.234**	-0.261**	-0.145**	52,895	-0.005	-0.318**	-0.188**
		[0.011]	[0.010]	[0.013]		[0.012]	[0.012]	[0.015]
1997	53,010	-0.227**	-0.258**	-0.140**	53,692	-0.003	-0.326**	-0.205**
		[0.011]	[0.010]	[0.013]		[0.012]	[0.012]	[0.015]
1998	52,468	-0.247**	-0.263**	-0.130**	53,477	0.016	-0.316**	-0.201**
		[0.011]	[0.011]	[0.013]		[0.012]	[0.012]	[0.015]
1999	50,649	-0.249**	-0.270**	-0.147**	51,455	0.017	-0.344**	-0.227**
		[0.011]	[0.011]	[0.013]		[0.012]	[0.012]	[0.015]
2000	49,630	-0.283**	-0.297**	-0.184**	50,597	0.043**	-0.359**	-0.242**
		[0.012]	[0.011]	[0.014]		[0.012]	[0.013]	[0.016]
2001	51,867	-0.267**	-0.300**	-0.176**	52,839	0.045**	-0.391**	-0.250**
		[0.012]	[0.011]	[0.015]		[0.012]	[0.013]	[0.016]
2002	56,549	-0.232**	-0.289**	-0.168**	56,958	0.072**	-0.362**	-0.236**
		[0.012]	[0.011]	[0.014]		[0.012]	[0.013]	[0.016]
2003	54,042	-0.217**	-0.239**	-0.082**	54,037	0.036**	-0.318**	-0.203**
		[0.012]	[0.012]	[0.014]		[0.012]	[0.014]	[0.017]
2004	53,120	-0.240**	-0.284**	-0.122**	53,481	0.067**	-0.319**	-0.219**
		[0.012]	[0.012]	[0.015]		[0.012]	[0.014]	[0.017]
2005	53,745	-0.228**	-0.280**	-0.110**	54,335	0.060**	-0.280**	-0.172**
		[0.013]	[0.012]	[0.014]		[0.012]	[0.014]	[0.017]
2006	53,463	-0.230**	-0.290**	-0.120**	54,415	0.072**	-0.282**	-0.154**
		[0.013]	[0.012]	[0.014]		[0.013]	[0.014]	[0.017]
2007	53,868	-0.244**	-0.279**	-0.119**	54,768	0.079**	-0.325**	-0.202**
		[0.013]	[0.012]	[0.015]		[0.012]	[0.014]	[0.017]
2008	53,742	-0.245**	-0.270**	-0.102**	54,081	0.069**	-0.289**	-0.150**
		[0.013]	[0.012]	[0.014]		[0.013]	[0.015]	[0.018]
2009	52,980	-0.223**	-0.281**	-0.120**	52,293	0.024+	-0.316**	-0.189**
		[0.013]	[0.012]	[0.015]		[0.013]	[0.015]	[0.018]
2010	50,219	-0.220**	-0.281**	-0.154**	49,650	0.026*	-0.328**	-0.226**
		[0.014]	[0.013]	[0.015]		[0.013]	[0.015]	[0.018]

Notes: The "no controls" specification reports FEM coefficients (θ_f and θ_m) from regressions with no other controls. The "standard" specification includes controls for years of schooling, potential experience (measured by age-schooling-5) and its square, and dummies for union coverage, large metropolitan area, full-time employment (usual hours worked are at least 35 hours), race/ethnicity (3) [Hipanic, Non-Hispanic Black, Non-Hispanic Other], marital status (2) [married with spouse present, married with spouse not present], region (8), industry/sector including the public sector (13), and occupation (5). The "expanded" specifications include all controls used in the "standard" specifications plus 10 additional occupational and industry controls. Robust standard errors are in brackets. **,* , + denote statistical significance at the 0.01, 0.05, and 0.1 levels, respectively.

Table 3 . Gender Composition Coefficients from Linear and Dummy Variable Models, Pooled Sample, 1996-2010

	Model 1	Model 2		
Specification	FEM	FEM25-49	FEM50-74	FEM75+
Women				
No Controls	-0.239**	0.060**	0.004	-0.096**
	[0.003]	[0.004]	[0.004]	[0.003]
Standard	-0.275**	-0.048**	-0.151**	-0.199**
	[0.003]	[0.003]	[0.003]	[0.003]
Expanded	-0.130**	0.000	-0.052**	-0.064**
	[0.004]	[0.003]	[0.003]	[0.004]
N	791,392			
Men				
No Controls	0.042**	0.172**	0.092**	-0.122**
	[0.003]	[0.002]	[0.002]	[0.003]
Standard	-0.324**	-0.006**	-0.131**	-0.230**
	[0.003]	[0.002]	[0.002]	[0.003]
Expanded	-0.202**	0.026**	-0.056**	-0.094**
	[0.004]	[0.002]	[0.003]	[0.003]
N	798,973			

Notes : In Model 1 the feminization variable is a continuous measure, while in Model 2 it is coded into three occupational female intensity dummies, where the reference group is FEM< 25%. Year dummies are included in all models. The "no controls", "standard" and "expanded" specifications are as defined in Table 2. Robust standard errors are in brackets. ** denotes statistical significance at the 0.01 level.

Table 4. Gender Composition Coefficient Sensitivity to Specification, Pooled Data, 1996-2010

Specification	Women	Men
1. No controls	-0.239**	0.042**
2. Base model (individual characteristics only)	-0.234**	-0.176**
3. Base model + 13 industry dummies	-0.272**	-0.103**
4. Base model + 5 occupation dummies	-0.232**	-0.367**
5. Standard model (base model + 5 occupation and 13 industry dummies)	-0.275**	-0.324**
6. Standard model + OPTD education & training	-0.202**	-0.274**
7. Standard model + Occupation tenure	-0.277**	-0.309**
8. Standard model + Occupation part-time	-0.171**	-0.179**
9. Standard model + O*NET computer	-0.248**	-0.289**
10. Standard model + OPTD education & training, Occupation tenure, Occupation part-time, and O*NET computer	-0.163**	-0.180**
11. Standard model + O*NET environment, hazards, physical, and strength	-0.285**	-0.335**
12. Standard model + Industry large firm and Industry union	-0.244**	-0.323**
13. Expanded model (standard model + all job characteristics)	-0.130**	-0.202**
14. Expanded model without O*NET physical	-0.122**	-0.199**
N	791,392	798,973

Notes: Coefficients shown are θ_f and θ_m. The "base" model excludes industry and occupation dummies. The "standard" and "expanded" specifications are as defined in Table 2. All models include year dummies. ** denotes statistical significance at the 0.01 level.

Table 5 . Gender Composition Coefficients among Different Worker Groups, Wage Level Equations, Pooled Sample, 1996-2010

	Women			Men		
Group	*N*	Standard	Expanded	*N*	Standard	Expanded
All workers	791,392	-0.275**	-0.130**	798,973	-0.324**	-0.202**
Age:						
16-29	172,905	-0.174**	-0.071**	182,661	-0.252**	-0.179**
30-39	199,161	-0.323**	-0.166**	212,754	-0.366**	-0.211**
40-49	213,583	-0.322**	-0.157**	206,742	-0.354**	-0.203**
50-59	151,275	-0.266**	-0.110**	143,850	-0.340**	-0.198**
60+	54,468	-0.137**	-0.069**	52,966	-0.252**	-0.170**
Marital Status:						
Married spouse present	448,417	-0.310**	-0.166**	500,766	-0.354**	-0.208**
Married spouse not present	167,753	-0.232**	-0.096**	99,459	-0.330**	-0.219**
Never married	175,222	-0.223**	-0.086**	198,748	-0.261**	-0.163**
Education (in years):						
0-11	54,434	-0.116**	-0.028+	81,570	-0.236**	-0.163**
12	243,230	-0.177**	-0.046**	258,298	-0.197**	-0.121**
13-15	246,025	-0.204**	-0.076**	216,221	-0.247**	-0.107**
16	168,283	-0.385**	-0.230**	159,886	-0.451**	-0.231**
>16	79,420	-0.452**	-0.250**	82,998	-0.345**	-0.080**
Race:						
Non-Hispanic						
White	599,149	-0.270**	-0.143**	602,775	-0.328**	-0.198**
Black	73,980	-0.247**	-0.071**	53,862	-0.242**	-0.131**
Other Race	43,774	-0.321**	-0.115**	44,336	-0.415**	-0.225**
Hispanic	74,489	-0.217**	-0.074**	98,000	-0.282**	-0.164**

Continues next page

Continued from previous page

	Women			Men		
Group	*N*	Standard	Expanded	*N*	Standard	Expanded
Sector:						
Private	623,467	-0.212**	-0.123**	676,325	-0.245**	-0.170**
Public	167,925	-0.461**	-0.141**	122,648	-0.469**	-0.259**
Union status:						
Nonunion	677,358	-0.256**	-0.139**	663,282	-0.279**	-0.185**
Union	114,034	-0.333**	-0.008	135,691	-0.375**	-0.174**
Working time status:						
Part-time	180,453	-0.031**	0.012	61,678	-0.139**	-0.038*
Full-time	610,939	-0.328**	-0.156**	737,295	-0.338**	-0.209**
Occupation:						
Managerial and Professional Specialty Occupations	299,369	-0.284**	-0.221**	246,940	-0.306**	-0.218**
Technical, Sales, and Administrative Support Occupations	291,370	-0.312**	-0.141**	140,777	-0.373**	-0.240**
Service Occupations	133,237	-0.153**	0.015	93,485	-0.186**	0.097**
Farming, Forestry, and Fishing Occupations	3,165	0.037	0.079	13,440	-0.321**	0.354**
Precision Production, Craft, and Repair Occupations	10,090	-0.378**	-0.137**	145,294	-0.143**	-0.201**
Operators, Fabricators, and Laborers	54,161	-0.218**	0.067**	159,037	-0.228**	-0.131**

Notes: Coefficients shown are θ_f and θ_m. The "standard" and "expanded" specifications are as defined in Table 2. All models include year dummies. **,*, + denote statistical significance at 0.01, 0.05, and 0.10 levels, respectively.

Table 6. Decomposition of the Gender Wage Gap, by Specification and Year

Specification	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Total log gap	0.232	0.234	0.226	0.235	0.226	0.226	0.214	0.202	0.199	0.194	0.191	0.191	0.195	0.192	0.182
Standard specification without FEM:															
1a. Unexplained	0.205	0.207	0.205	0.211	0.206	0.211	0.206	0.183	0.190	0.190	0.181	0.187	0.189	0.181	0.175
1b. Total explained	0.027	0.027	0.022	0.024	0.020	0.016	0.007	0.019	0.009	0.004	0.010	0.004	0.007	0.011	0.006
Standard specification:															
2a. Unexplained	0.139	0.142	0.140	0.146	0.136	0.140	0.139	0.127	0.128	0.132	0.123	0.128	0.132	0.121	0.117
2b. Total explained	0.093	0.091	0.086	0.089	0.090	0.086	0.075	0.075	0.071	0.062	0.068	0.064	0.063	0.071	0.065
2c. Explained due to FEM	0.108	0.107	0.108	0.110	0.118	0.121	0.115	0.099	0.111	0.104	0.105	0.109	0.101	0.106	0.106
Expanded specifications:															
3a. Unexplained	0.143	0.145	0.143	0.148	0.138	0.143	0.141	0.134	0.135	0.138	0.129	0.132	0.137	0.128	0.124
3b. Total explained	0.089	0.089	0.083	0.087	0.088	0.083	0.073	0.068	0.064	0.056	0.062	0.059	0.058	0.064	0.058
3c. Explained due to FEM	0.052	0.054	0.054	0.058	0.069	0.066	0.064	0.039	0.052	0.042	0.040	0.048	0.036	0.045	0.054
3d. Explained due to all job characteristics	0.031	0.028	0.026	0.024	0.016	0.018	0.022	0.020	0.015	0.014	0.020	0.013	0.021	0.018	0.007
3e. Explained due to selected job characteristics															
Education & Training	0.002	0.001	0.001	0.001	0.001	0.000	-0.001	0.000	0.000	-0.001	0.000	-0.002	0.000	-0.001	-0.002
Computers	-0.006	-0.007	-0.007	-0.006	-0.006	-0.005	-0.005	-0.005	-0.006	-0.006	-0.005	-0.006	-0.005	-0.003	-0.004
Physical	-0.014	-0.013	-0.009	-0.013	-0.013	-0.014	-0.014	-0.013	-0.010	-0.012	-0.014	-0.013	-0.008	-0.009	-0.010
Part-time	0.024	0.025	0.021	0.025	0.020	0.027	0.029	0.019	0.019	0.018	0.022	0.017	0.018	0.020	0.013

Notes: The "standard" and "expanded" specifications are as defined in Table 2. Decompositions are performed using the *oaxaca* command in Stata 12 with the pooled option; that is, a twofold decomposition using the coefficients from a pooled model over both groups as the reference coefficients. A group indicator is also used in the pooled model to avoid potential distortion caused by the residual group difference spilling over into the slope parameters of the pooled model, as recommended by Jann (2008).

Table 7. FEM Coefficients from Estimates with Panel Data from the CPS-MORG and NLSY79, Wage Level and Wage Change Equations

	Women			Men		
Dataset (Years) / Sample / Model	Standard	Expanded	Expanded Plus	Standard	Expanded	Expanded Plus
CPS-MORG (1996-2010) / Occupation and Industry Changers						
OLS	-0.308**	-0.232**	-	-0.298**	-0.261**	-
	[0.009]	[0.012]		[0.011]	[0.014]	
FD	-0.050**	-0.023+	-	-0.043**	-0.004	-
	[0.010]	[0.013]		[0.011]	[0.014]	
Number of Observations	69,778			76,300		
Number of Individuals	34,889			38,150		
NLSY79 (1993-2010)						
OLS	-0.291**	-0.160**	-0.152**	-0.349**	-0.235**	-0.246**
	[0.018]	[0.021]	[0.020]	[0.022]	[0.026]	[0.026]
FD	-0.112**	-0.036	-0.036	-0.073**	-0.011	-0.001
	[0.021]	[0.027]	[0.026]	[0.027]	[0.034]	[0.034]
FE	-0.126**	-0.045*	-0.043*	-0.063**	-0.015	-0.009
	[0.018]	[0.022]	[0.022]	[0.022]	[0.028]	[0.028]
Number of Observations	16,772			16,250		
Number of Individuals	2,586			2,393		

Notes: In the CPS-MORG wage change is measured over one-year intervals, so that these estimates are not directly comparable with those from the NLSY79, but see text for an operationalization. In using NLSY79 data for the FD regressions we dropped the 1993 round in order to have a consistent measure of wage change over two-year intervals. In these first differenced regressions there are 10,987 observations for women and 10,831 observations for men, fewer than for the OLS and FE regressions (16,772 women and 16,250 men). The "standard" and "expanded" specifications are as defined in Table 2. In the "expanded plus" specification the measure of potential experience is replaced by actual labor market experience, tenure, and age variables. This latter specification also controls for age- and education-adjusted ASVAB scores. **,*, + denote statistical significance at 0.01, 0.05, and 0.10 levels, respectively.

Table 8. Robustness Checks for the FEM Coefficients from the CPS-MORG Wage Change Equations

Sample/Cohort/Model			Women Standard	Women Expanded	Men Standard	Men Expanded
Occupation and Industry Changers	Aging Cohort	FD	-0.026	0.001	-0.049*	0.001
			[0.018]	[0.023]	[0.021]	[0.027]
	Number of Observations		18,848		20,154	
	Number of Individuals		9,424		10,077	
	Younger Aging Cohort	FD	-0.087**	-0.068**	-0.064**	-0.043
			[0.021]	[0.026]	[0.022]	[0.028]
	Number of Observations		15,726		20,154	
	Number of Individuals		7,863		10,077	
	Older Aging Cohort	FD	-0.035	-0.015	-0.008	0.022
			[0.024]	[0.032]	[0.024]	[0.032]
	Number of Observations		96,360		89,194	
	Number of Individuals		48,180		44,597	

Notes: See the text for information how the various samples are configured. **,* denote statistical significance at 0.01 and 0.05 levels, respectively. The "standard" and "expanded" specifications are as defined in Table 2.